

PRE-APPEAL BRIEF REQUEST FOR REVIEW		Docket Number (Optional) 0113744.00124US2
	Application Number 10/803,831-Conf. #5913	Filed March 18, 2004
	First Named Inventor John R. FRANK	
	Art Unit 2168	Examiner O. Omosewo

Applicant requests review of the final rejection in the above-identified application. No amendments are being filed with this request.

This request is being filed with a notice of appeal.

The review is requested for the reason(s) stated on the attached sheet(s).

Note: No more than five (5) pages may be provided.

I am the

- applicant /inventor.
 assignee of record of the entire interest.
 See 37 CFR 3.71. Statement under 37 CFR 3.73(b)
 is enclosed. (Form PTO/SB/96)

- attorney or agent of record.

Registration number 56,451



Signature

Oliver Strimpel

Typed or printed name

- attorney or agent acting under 37 CFR 1.34.

Registration number if acting under 37 CFR 1.34. _____

(617) 526-6000

Telephone number

January 10, 2008

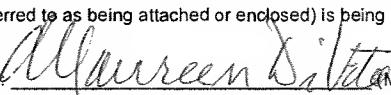
Date

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required.
 Submit multiple forms if more than one signature is required, see below*.

*Total of 1 forms are submitted.

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4).

Dated: January 10, 2008

Signature 
(Maureen DiVito)

The Examiner rejected claim 1 under 35 U.S.C. § 102(b) as being anticipated by Japanese Publication No. 2002/0031269 to Toshikazu Fukushima (Toshikazu). But contrary to what the Examiner appears to believe, Toshikazu fails to disclose two key features of the claim. More specifically, Toshikazu fails to disclose “in a large corpus, identifying geo-textual correlations among readings of the toponyms within the plurality of toponyms” and he fails to disclose “using the identified geo-textual correlations to generate a value for a confidence that the selected toponym refers to a corresponding geographic location.” Rather than identifying geo-textual correlations, Toshikazu resolves ambiguities in the meanings names simply by looking for the presence of “co-occurring words” which he identifies through a look-up table. We explain the differences in more detail below.

Before we look at the meaning of “geo-textual correlations,” it is important to understand what a reading of a toponym is. A reading of a toponym is a geographical location with which the toponym is associated, such as a latitude-longitude or an area. For example, a reading of “Paris” is the geographic region associated with Paris, France. Many toponyms have more than one reading. In the case of “Paris,” it also could be the geographic region associated with the town of Paris, Texas.

According to the instant specification, there is a statistical property of documents that reveals a relationship between readings of toponyms and their relative locations within a document’s text. The present specification refers to that property as a geo-textual correlation. In general terms, the applicant has observed that toponyms that have readings that are close to each other in geographical space are more likely to be close to each other in the text of a document. The specification explains this in greater detail:

A technical advance is achieved in the art by exploiting knowledge of a hitherto unobserved statistical property of documents, namely geo-textual correlation. By inspecting large corpora, we have found that there is a high degree of spatial correlation in geographic references that are in textual proximity. This applies not only to points that are nearby (such as Madison and Milwaukee), but also to geographic entities that enclose or are enclosed by regions (Madison and Wisconsin, for example). More specifically, if the textual distance between names N and M is small, and if N has a reading P (i.e., N is associated with P or N means P) and M has a reading Q, then the physical distance between P and Q is likely to be lower than would be expected randomly. Conversely, if P and Q are close geographically, then their names N and M are more likely to appear together in texts than would be expected randomly. This correlation between geographic and textual distance is considered in estimating of the confidence $c(N,P)$ that a name N refers to a particular point P. (page 7, line 17-28, emphasis added)

Thus, for example, since Madison and Milwaukee are geographically close to each other (i.e., they have readings that are close to each other), the words “Madison” and “Milwaukee” are statistically likely to appear close to each other within the text of a document. Conversely, if the words Madison and Milwaukee often appear close to each other within documents of a large corpus of documents, then they are statistically likely to have readings that are close to each other. **These geo-textual correlations are identified by statistically analyzing a large number of documents. Indeed, the phrase “geo-textual correlations” implies conducting such a statistical analysis.**

Claim 1 recites “in a large corpus, identifying geo-textual correlations among readings of ... toponyms.” The Examiner asserts that Toshikazu discloses this, and directs our attention to two passages (i.e., ¶79 and ¶80). The first passage reads as follows:

Incidentally, there are varieties of calculation methods in terms of appearance frequency information of co-occurring words in plural texts. In FIG. 9, for example, the location of “Chuo-ku” in the text 19 is not to be specified by referring to the referring link text 17, in which both co-occurring words “Tokyo” and “Osaka” appear. Consequently, according to the process (D), the analysis is performed referring to plural referring link texts. Additionally, even a linked text(s) is subject to the reference. Referring to the linked text 20 as well as the referring link texts 17 and 18, it is turned out that the co-occurring words “Tokyo”, “Kinki-Area”, and “Kyoto” appear in the texts once respectively, and “Osaka” appears three times. Thus “Chuo-ku” can be taken as “Chuo-ku” in Osaka in recognition of that “Osaka” makes the most of appearance. (¶ [0079], emphasis added)

But this passage has nothing to do with identifying geo-textual correlations in a large corpus of documents. The term “geo-textual correlations” implies a statistical analysis of a corpus of documents – typically a large corpus to make the statistical observations meaningful. Toshikazu does not perform any kind of statistical analysis of a corpus of documents to generate any correlations. Rather, Toshikazu’s method involves resolving the ambiguity in the meaning of a location name by counting the frequencies of the appearance of co-occurring words, which he retrieves from a “named entity dictionary.” (see Figure 7). His “named entity dictionary” identifies what are believed to be associations and that information was generated in a different way from how geo-textual correlations are determined. The named entity dictionary is assembled by methods referred to in Toshikazu’s background section, which include using pre-existing databases populated by various methods of extracting proper nouns. For example, Toshikazu refers to “a proper noun extracting means for extracting candidates for the proper noun from the text obtained by the database accessing means with reference to patterns of proper

nouns prepared in advance." (¶ [0008]) Toshikazu further describes his named entity dictionary as follows:

The named entity dictionary 33 stores a dictionary for identifying the candidate named entities. ...the named entity dictionary contains potential categories 41, such as "location name", "personal name", and "organization name", for each term of the named entities (sic) 40. And further, the dictionary stores a co-occurring word list 42 for each category. It is preferable that not only the co-occurring words but also their positional condition (for example, "collocating with the named entity", etc.) is added to the co-occurring word list 42. (¶ [0063])

But none of these methods, or anything else mentioned in the above paragraph, discloses identifying geo-textual correlations in a large corpus. He simply uses information that is stored in his pre-existing databases and that was acquired through other means.

The other passage to which the Examiner directed our attention reads as follows:

In the above method, the co-occurring word that appears most frequently in the plural texts has priority. On the other hand, there is another method in which the co-occurring word that appears in the most numbers of referring link and linked texts has priority. Referring to FIG. 9, for example, the co-occurring word "Osaka" appears in the three texts, 17, 18, and 20, while "Kinki-Area" and "Kyoto" make their appearance in only the text 18. Thereby "Osaka" is regarded as the co-occurring word appearing in the most numbers of texts, and thereby used as a clue to resolve the ambiguity in the candidate named entity. (¶ [0080])

This passage discloses a variant of Toshikazu's toponym ambiguity-resolving method that exploits the hyperlink structure of his texts. As in the previous passage, the method involves looking up the toponym to be resolved in a pre-compiled dictionary of proper names to acquire a list of co-occurring words associated with that toponym. Then, the co-occurring word that appears in the greatest number of pages linked to the toponym to be resolved is selected to resolve the ambiguity. As before, there is no reference to "in a large corpus, identifying geo-textual correlations among readings of the toponyms within the plurality of toponyms," as required by the claim. Furthermore, we are unable to find even a hint of such a reference anywhere within Toshikazu.

As mentioned above, Toshikazu is deficient in another very important way. Toshikazu also fails to disclose "using the identified geo-textual correlations to generate a value for a confidence that [a] selected toponym refers to a corresponding geographic location." (emphasis added). Though the Examiner appears to believe otherwise, Toshikazu says nothing whatsoever about confidence values. Indeed, the word "confidence" does not even appear in the Toshikazu

application. Toshikazu simply resolves ambiguities by picking the “best” alternative which is the one that has the highest number of co-occurring words in linking documents.

The Examiner also rejected claim 10 under 35 U.S.C. § 102(b) as being anticipated by Toshikazu. But contrary to what the Examiner states, Toshikazu does not disclose:

a document that includes a plurality of toponyms for which there is a corresponding plurality of (toponym,place) pairs, there being associated with each (toponym,place) pair of said plurality of (toponym,place) pairs a corresponding value for a confidence that the toponym of that (toponym,place) pair refers to the place of that (toponym,place) pair...

as required by the claim. (emphasis added). The Examiner appears to believe that ¶’s 79-81 disclose this. But these paragraphs (two of which are presented and discussed above) involve documents having a set of names for which Toshikazu looks up “co-occurring words,” in a table. His table (see, e.g., Figure 7 and ¶ [0063]) includes a list of named entities (Fig. 7, 40), and for each named entity, a list of co-occurring words (Fig. 7, 42). But the table contains no (toponym,place) pairs, and more importantly it does not include any confidence values that the co-occurring words refer to a particular place, as required by the claim. We were also unable to find any mention of the (toponym,place) structure or of confidence values anywhere else within Toshikazu.

We note that claim 10 also recites “boosting the value of the confidence for a selected (toponym,place) pair.” This feature is completely absent from the teachings of Toshikazu. As discussed above, Toshikazu does not compute or store confidence values in his named entity table, or anywhere else for that matter. So, he has no confidence values to boost.

The Examiner further rejected claim 18 under 35 U.S.C. § 102(b) as being anticipated by Toshikazu. Claim 18 requires “...identifying a plurality of (toponym,place) pairs that is associated with the selected document, and for each identified (toponym,place) pair, obtaining and using a value for a confidence that the toponym of the (toponym,place) pair refers to the place.” (emphasis added) As discussed above for claim 10, Toshikazu makes no mention of (toponym,place) pairs, nor does he disclose obtaining and using confidence values that a toponym refers to a place.

For the reasons discussed above, Applicant believes that claims 1, 10, and 18, and dependent claims 2-9 and 11-17 are not anticipated by Toshikazu and therefore asks that this application be allowed to issue.